

Repeated multimarket contact with observation errors

Atsushi Iwasaki¹, Tadashi Sekiguchi², Shun Yamamoto³, and Makoto Yokoo³

¹ University of Electro-Communications, iwasaki@is.uec.ac.jp

² Kyoto University, sekiguchi@kier.kyoto-u.ac.jp

³ Kyushu University, {syamamoto@agent.,yokoo@}inf.kyushu-u.ac.jp

Abstract. This paper analyzes repeated multimarket contact with observation errors where two players operate in multiple markets simultaneously. Multimarket contact has received much attention in economics, management, and so on. Despite vast empirical studies that examine whether multimarket contact fosters cooperation or collusion, little is theoretically known as to how players behave in an equilibrium when each player receives a noisy and different observation or signal indicating other firms' actions (*private* monitoring). To the best of our knowledge, we are the first to construct a strategy designed for multiple markets whose per-market equilibrium payoffs exceed one for a single market, in our setting. We first construct an entirely novel strategy whose behavior is specified by a non-linear function of the signal configurations. We then show that the per-market equilibrium payoff improves when the number of markets is sufficiently large.

1 Introduction

This paper analyzes repeated multimarket contact with observation errors where two players operate in multiple markets simultaneously. A firm, e.g., Uber, provides its taxi service in multiple distinct markets (areas) and determines its price or allocation in each area, facing an oligopolistic competition, which is often modeled as a prisoners' dilemma (PD). To improve profits, it is inevitably helpful to realize how the firm's rival should behave in an equilibrium. Alternatively, it is pointed out that tacit collusion among firms is likely to occur [7]. It is also desirable for a regulatory agency to theoretically understand the extent of the profits firms earn by collusion.

However, despite vast empirical studies [23] that have examined whether multimarket contact fosters cooperation or collusion, little is theoretically known as to how players behave in an equilibrium when each player receives a noisy observation or signal of other firms' actions. Without noisy observation, i.e., under *perfect* monitoring, where each player can observe his opponents' actions, there exists no strategy designed for multiple markets whose per-market equilibrium payoff exceeds one for a single market [3]. With noisy observation, one exception is a case where players do share common information, i.e., *public* monitoring where all players always observe a noisy, but common signal. A generalization of

trigger strategies attains greater per-market equilibrium payoffs than the single-market equilibrium, assuming a public randomization device [14].

In contrast, this paper considers a different, but realistic noisy situation where players do not share common information on each other’s past history, i.e., *private* monitoring where each player may observe a different signal. For example, although a firm cannot directly observe its rival’s action, e.g., prices, it can observe a noisy signal, e.g., its rival’s sales amounts. Analytical studies on this class of games have not been very successful. Though the repeated PD with observation errors has been extensively studied, most papers assume public monitoring in the literature of economics [17]. This is because finding equilibria in such games has been considered to be extremely hard. Indeed, it requires very complicated statistical inferences to estimate the history a player reaches at a period and to compute the continuation payoff from the period on [13]. Notably, a *belief-free* approach has successfully established a general characterization where an equilibrium strategy is constructed so that a player’s belief (about her opponent) does not matter [11,10]. However, it is not obvious whether the belief-free approach is helpful in examining the effects of multimarket contact, because we want to deal with any number of markets. Its tractability may be lost if the number of markets increases, so that the number of available actions exponentially increases.

The goal of this paper is to answer the following question: under multimarket contact with private monitoring, can we find a particular class of strategies which can sustain a better outcome than an equilibrium strategy for a single market? For a benchmark, we focus on a strategy found by Ely and Välimäki [11] that attains the optimal payoff among belief-free equilibria in PD. Figure 1 illustrates the strategy, which we call EV, as a variant of the well-known tit-for-tat strategy.⁴

To the best of our knowledge, we are the first to construct a strategy designed for multiple markets whose per-market equilibrium payoffs exceed one for a single market. First, we construct an entirely novel strategy whose behavior is specified by a nonlinear function of the signal configurations. Precisely, a player chooses her action at a period according to in which markets she receives bad signals at the previous period. We call this class of the strategies *nonlinear transition, partial defection* (NTPD). Then, we show that the per-market equilibrium payoff improves when the number of markets is sufficiently large via the theoretical and numerical analysis.

In the literature of computer science, AI, and multi-agent systems, there are many streams associated with repeated games [6]: the complexity of equilibrium computation [16,5,2], multi-agent learning [4,8,19], partially observable stochastic games (POSGs) [12,9,21,18,22], and so on. Among them, POSGs are the most relevant to repeated games with private monitoring because they can be considered as a special case of POSGs. However, POSGs often impose partial

⁴ Here, g or b is a private signal, which is a noisy observation of opponent actions C or D . ε_R or ε_P represents the transition probability between states. We omit the remaining ones.

observability on an opponent’s strategy (behavior rule) and not on opponent’s past actions [18,22]. They estimate an optimal (best reply) strategy against an unknown strategy (not always fixed) from perfectly observable actions (perfect monitoring). In contrast, we verify whether a given strategy profile is a *mutual* best reply after any history, i.e., finding an equilibrium, with partially observable actions (private monitoring). Thus, this paper also addresses understanding the gap between POSGs and repeated games with private monitoring in economics.

In fact, very few existing works have addressed verifying an equilibrium. Hansen, Bernstein, and Zilberstein [12] develop an algorithm that iteratively eliminates dominated strategies. However, just eliminating dominated strategies is not sufficient to find an equilibrium. Also, the algorithm is not applicable to an infinitely repeated game. Doshi and Gmytrasiewicz [9] investigate the computational complexity of achieving equilibria in interactive POMDPs.

Among an enormous number of studies on repeated games in economics, an important topic has been validity of the folk theorem. Most of them assume perfect or public monitoring (Please consult the textbook [17]). In case of private monitoring, a recent paper [20] establishes a general folk theorem. However, the result is irrelevant to our analysis because the equilibrium strategies are excessively complicated and require nearly complete patience of the players. Specializing in multimarket contact, we rather show that the NTPD strategy forms a highly cooperative equilibrium and only requires the players to be mildly patient.

2 Model

Two players play M PDs simultaneously in each period. In each PD, each player chooses either C (cooperation) or D (defection). This is regarded as a model of oligopolistic competition, where C is an action increasing the total payoffs (for instance, in the case of price competition, charging a collusive high price), and D is a non-cooperative one (like a price cut). The players can choose different actions over the M PDs, so that each player’s action set in each period is $\{C, D\}^M$.

Each player cannot directly observe the other player’s actions, but receives an imperfect signal about them. In each PD, each player receives either a good signal g or a bad signal b . We assume that each player receives his signals individually, and cannot observe the other player’s signals (private monitoring). The pair of signals they privately receive in each PD is stochastic, following a common symmetric probability distribution that depends entirely on the action pair of that PD. We denote it by $o(\omega_1, \omega_2 | a_1, a_2)$, where $(\omega_1, \omega_2) \in \{g, b\}^2$ and $(a_1, a_2) \in \{C, D\}^2$. We assume that the signals across the M PDs are independent, though the signals of a given PD may be correlated across the players. We also assume that the signal distributions are described by one parameter. There exists $p \in$

$(1/2, 1)$ such that for any i , any ω_j ($j \neq i$) and any $a \in \{C, D\}^2$,

$$\sum_{\omega_i \in \{g, b\}} o(\omega_i, \omega_j | a) = \begin{cases} p & \text{if } (a_i, \omega_j) \in \{(C, g), (D, b)\}, \\ 1 - p & \text{otherwise.} \end{cases}$$

The marginal distribution of an individual signal in a given PD is such that the *right* signal ($\omega_j = g$ if $a_i = C$, and $\omega_j = b$ if $a_i = D$) is received with probability p . We let $s = 1 - p$, which is the probability of an *error*. The assumption is consistent with *conditionally independent* monitoring, which is a representative monitoring structure. Formally, a signal distribution is conditionally independent if $o(\omega_i, \omega_j | a) = o(\omega_i | a)o(\omega_j | a)$ for all ω_i, ω_j , and a .

In each PD, player i 's payoff depends only on his action and the signal of that PD. The payoff function is common to all PDs, denoted by $\pi_i(a_i, \omega_i)$. We are more interested in the expected payoff function:

$$g_i(a_1, a_2) = \sum_{(\omega_1, \omega_2)} \pi_i(a_i, \omega_i) o(\omega_1, \omega_2 | a_1, a_2).$$

We assume that their expected payoff functions are represented by the following payoff matrix:

	C	D
C	$1, 1$	$-y, 1 + x$
D	$1 + x, -y$	$0, 0$

We assume $x > 0$, $y > 0$ and $1 > x - y$, so that it indeed represents a PD.

All M PDs are played infinitely, in periods $t = 0, 1, 2, \dots$. Player i 's *private history* at the beginning of period $t \geq 1$ is an element of $H_i^t \equiv [\{C, D\}^M \times \{g, b\}^M]^t$. Let H_i^0 be an arbitrary singleton, and let $H_i = \cup_{t \geq 0} H_i^t$ be the set of player i 's all private histories. Player i 's strategy of this repeated game is a mapping from H_i to the set of all probability distributions over $\{C, D\}^M$. That is, we allow randomized strategies. If the actual play of the repeated game is such that the action pair $(a_1^m(t), a_2^m(t))$ is played in the m -th PD in period t for each m and t , player i 's normalized average payoff is

$$(1 - \delta) \sum_{t=0}^{\infty} \delta^t \sum_{m=1}^M g_i(a_1^m(t), a_2^m(t)), \quad (1)$$

where $\delta \in (0, 1)$ is their common discount factor. The average payoff of any strategy pair is the expected value of Eq. 1, where the expectation is taken with respect to the players' randomizations and the monitoring structure.

The standard solution concept for repeated games with imperfect monitoring is *sequential equilibrium* [15], but here we focus on a special class called *belief-free equilibria* [10].

Definition 1 (Belief-free equilibrium). A strategy pair is a *belief-free equilibrium* if for any $t \geq 0$, $h_1^t \in H_1^t$ and $h_2^t \in H_2^t$, each player i 's continuation strategy given h_i^t is optimal against player j 's continuation strategy given h_j^t .

An important property of the belief-free equilibria is that, while player i given her private history should, in principle, optimize her continuation payoff against her belief about player j 's history (and hence his continuation strategy), her continuation strategy is optimal even if she were to know j 's history with certainty.⁵ In other words, the players playing a belief-free equilibrium need not compute their beliefs in the course of play. When a strategy pair is represented by finite-state automaton strategies, as will be the case in subsequent analysis, it is a belief-free equilibrium if any player's continuation strategy (behavior expanded from the automaton) starting from any state is a best response (optimal) against the other player's continuation strategy starting from any state. Note that we never restrict the other's possible strategy space, which includes strategies with an infinite number of states.

Suppose both players employ a common strategy represented by a two-state automaton with state space $\{R, P\}$. Let $V_{s_1 s_2}$, where $s_1 \in \{R, P\}$ and $s_2 \in \{R, P\}$, be player 1's continuation payoff when (i) player 2 is currently at s_2 and then follows the automaton, and (ii) player 1 always plays the action prescribed at state s_1 at any subsequent history. The strategy pair is a belief-free equilibrium if and only if there exist V_R and V_P such that

$$V_{RR} = V_{PR} = V_R, \quad V_{RP} = V_{PP} = V_P, \quad (2)$$

and that V_{s_2} ($s_2 \in \{R, P\}$) is player 1's best response payoff against player 2's continuation strategy when he is at state s_2 . To see this, note that by Eq. 2, player 1 at any history is indifferent between her continuation strategy at state R and that at state P irrespective of her belief about player 2's state. Since the second condition implies that both continuation strategies give her best response payoff at any history, the conditions for belief-free equilibrium are all satisfied.

Let us explain the EV strategy [11] depicted in Figure 1. A solid line denotes a deterministic transition and a dashed line denotes a probabilistic transition, though, for simplicity, we omit some state transition. EV is a representative two-state automaton strategy that forms a belief-free equilibrium under repeated games with private monitoring and attains the highest average payoff among belief-free equilibria in PD. It is parameterized by two numbers, $\varepsilon_R \in [0, 1]$ and $\varepsilon_P \in [0, 1]$. A player first cooperates at state R , but after observing a bad signal, she punishes (defects) at the next period with probability ε_R , or keep cooperation with $1 - \varepsilon_R$. Likewise, after she defects at P , if she observes a good signal, she returns cooperation with ε_P , or keep defection with $1 - \varepsilon_P$.

Proposition 1. *There exist $\varepsilon_R \in [0, 1]$ and $\varepsilon_P \in [0, 1]$ such that the EV strategy pair is a belief-free equilibrium if*

$$\delta[2p - 1 - (1 - p)(x + y) + \max\{x, y\}] \geq \max\{x, y\}. \quad (3)$$

The average payoff starting from state R is

$$V_R = 1 - \frac{(1 - p)x}{2p - 1}. \quad (4)$$

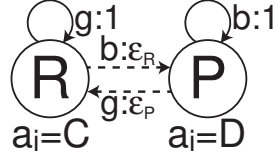


Fig. 1: EV strategy

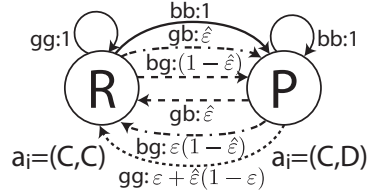


Fig. 2: NTPD strategy

What happens if there are $M(\geq 2)$ PDs, in comparison with the case of one PD? If EV forms an equilibrium, it is always an equilibrium to play it in each PD independently. Obviously, the payoff of this equilibrium is M times the EV equilibrium payoff. Under this equilibrium, a player's actions in all PDs can be quite different, depending on the histories of individual PDs. Thus, the corresponding automaton has 2^M states. The transition probabilities from one state to another state depends linearly on the number of bad signals. We now ask how this feature of EV can be modified so that the equilibrium profit increases with two-state automaton strategies.

3 Nonlinear transition, partial defection strategy

The goal of the analysis of this section is to find a particular class of strategies which can sustain a better payoff outcome. The strategies in this class are representable by two-state automata, and have the following two features: (i) the state transition probabilities are not linear in the number of bad signals, and (ii) defection in some PDs (*partial defection*) is prescribed at state P . Let us define the proposed class of strategies, which we call the *nonlinear transition, partial defection* (NTPD) strategy, given M PDs:

Definition 2 (NTPD strategy). *An NTPD strategy for $M(\geq 2)$ PDs is a two-state automaton strategy, parameterized by an integer M_A such that $1 \leq M_A < M$ and two numbers $\varepsilon \in [0, 1]$ and $\hat{\varepsilon} \in [0, 1]$. Let $A = \{1, 2, \dots, M_A\}$ and $B = \{M_A + 1, M_A + 2, \dots, M\}$.*

- The state space is $\{R, P\}$, and R is the initial state.
- At state R , the player is prescribed to choose C in all PDs. At state P , she is prescribed to choose C in all PDs in A and D in all PDs in B .
- Suppose the current state is R and k is an integer between 0 and $M_B = M - M_A$. Then
 1. if b is observed among **all PDs in A** and there are k bad signals among the PDs in B , then the state shifts to P with probability $1 - (M_B - k)\hat{\varepsilon}$ (and stays at R with the remaining probability).
 2. if g is observed among **some PD in A** and there are k bad signals among the PDs in B , then the state shifts to P with probability $k\hat{\varepsilon}$ (and stays R with the remaining probability).

⁵ We refer to player i or 1 as *her* and to player j or 2 as *him* throughout this paper.

- Suppose the current state is P and k is an integer between 0 and M_A . Then
 1. if g is observed among **all PDs in B** and there are k bad signals among the PDs in A , then the state shifts to R with probability $\varepsilon + \hat{\varepsilon}\{(1-\varepsilon)M_A - k\}$ (and stays P with the remaining probability).
 2. if b is observed among **some PD in B** and there are k bad signals among the PDs in A , then the state shifts to R with probability $(M_A - k)\hat{\varepsilon}$ (and stays P with the remaining probability).

Figure 2 illustrates NTPD for two PDs in the same manner as Figure 1. Let us explain how we construct this strategy. A player cooperates in all PDs in A at state P . Then, she always cooperates in all PDs in A regardless which state she is in. The transition probabilities from R to P distinguish signals from A with from B . The increase of transition probabilities is constant for the number of bad signals from PDs in A . If she observes at least one good signal from A , it is zero, otherwise, $1 - M_B\hat{\varepsilon}$. The transition probabilities further increase by $\hat{\varepsilon}$ in the number of bad signals k in B . For k bad signals from B , the transition probability from R to P is specified as $1 - (M_B - k)\hat{\varepsilon}$ if she observes some g in A , or $k\hat{\varepsilon}$ otherwise.

In a similar way, the transition probabilities from P to R are specified. Their increase is constant for the number of bad signals from PDs in B . If she observes at least one bad signal from B , it is zero, otherwise, $\varepsilon - \hat{\varepsilon}\varepsilon M_A$. The transition probabilities decrease by $\hat{\varepsilon}$ in the number of bad signals k in A . For k bad signals from A , the transition probability from P to R is specified as $(M_A - k)\hat{\varepsilon}$ if she observes some b in B , or $\varepsilon + \hat{\varepsilon}\{(1-\varepsilon)M_A - k\}$ otherwise. We here mix $1 - k\hat{\varepsilon}$ with $(M_A - k)\hat{\varepsilon}$ by the last parameter ε .⁶ In fact, if she observes M_A bad signals in A and M_B good signals in B , she transits to R with probability $\varepsilon - \hat{\varepsilon}\varepsilon M_A$.

3.1 Two PDs

Let us first analyze the case of two PDs.

Theorem 1 (NTPD for two PDs). *Fix $M = 2$ and $M_A = 1$. There exist ε and $\hat{\varepsilon}$ such that the NTPD strategy pair is a belief-free equilibrium if*

$$\delta[2p - 1 - (1-p)y + p \max\{x, y\}] \geq x + \max\{x, y\}. \quad (5)$$

The average payoff starting from R is

$$V_R = 2 - \frac{\delta(1-p)\{2p - 1 - px - (1-p)y\}}{(2p-1)(1-\delta p)}. \quad (6)$$

If the coefficient of δ in Eq. 5 is nonpositive, no δ satisfies it. Or if it does not hold at $\delta = 1$, then no δ satisfies it. Furthermore, if $x \geq 1$, Eq. 5 does not hold under any δ and p . Otherwise, it holds for all sufficiently large δ and p . A similar argument is applied to the case of M PDs.

⁶ To solve a system of value equations, we use only two parameters to determine the transition probabilities.

Proof. Suppose $M = 2$ and Eq. 5 hold. Define Eq. 6,

$$V_P = 1 + \frac{px + (1-p)y}{2p-1},$$

$$\varepsilon = \frac{(1-\delta p)y}{\delta\{2p-1-(1-p)y\} - x}, \quad \hat{\varepsilon} = \frac{(1-\delta p)x}{\delta\{2p-1-px-(1-p)y\}}.$$

From Eq. 5, we obtain

$$V_R > V_P, \quad 0 < \varepsilon \leq 1, \quad 0 < \hat{\varepsilon} \leq \frac{1}{2}. \quad (7)$$

Hence the NTPD strategy with ε and $\hat{\varepsilon}$ defined above, together with $M_A = 1$, is well-defined.

Some calculations verify that

$$V_{RR} = (1-\delta)2 + \delta V_{RP} + \delta p(V_{RR} - V_{RP}), \quad (8)$$

$$V_{PR} = (1-\delta)(2+x) + \delta V_{PP} + \delta\{p-(2p-1)\hat{\varepsilon}\}(V_{PR} - V_{PP}), \quad (9)$$

$$V_{RP} = (1-\delta)(1-y) + \delta V_{RP} + \delta p(\varepsilon + \hat{\varepsilon} - \varepsilon\hat{\varepsilon})(V_{RR} - V_{RP}), \quad (10)$$

$$V_{PP} = 1 - \delta + \delta V_{PP} + \delta\{p\hat{\varepsilon} + (1-p)\varepsilon(1-\hat{\varepsilon})\}(V_{PR} - V_{PP}). \quad (11)$$

Solving these, we obtain Eq. 2 for V_R and V_P defined above. These imply that (i) a player is indifferent between starting with state R and then conforming to NTPD and starting with state P and then conforming to NTPD, if the other player starts with either state R or P and then conforming to NTPD, (ii) a player's continuation payoff at one of the states when the other player is at $s \in \{R, P\}$ is V_s .

Let V_s^{DC} be a player's payoff when he selects D in the first PD and C in the second, and then conforms to NTPD when the other player's current state is $s \in \{R, P\}$. Similarly, let V_s^{DD} be a player's payoff when he selects D in both PDs, and then conforms to NTPD when the other player's current state is $s \in \{R, P\}$. The proof is complete if we show that $V_s \geq \max\{V_s^{DC}, V_s^{DD}\}$ for any s , i.e., each player at any state has no incentive to deviate from the action at that state. It is easy to verify that

$$V_R^{DC} = (1-\delta)(2+x) + \delta\left[\{1-p+(2p-1)\hat{\varepsilon}\}V_R + \{p-(2p-1)\hat{\varepsilon}\}V_P\right], \quad (12)$$

$$V_R^{DD} = (1-\delta)(2+2x) + \delta\{(1-p)V_R + pV_P\}, \quad (13)$$

$$V_P^{DC} = (1-\delta)(1+x-y) + \delta\left[\{\hat{\varepsilon} + p(\varepsilon - \hat{\varepsilon} - \varepsilon\hat{\varepsilon})\}V_R + \{1 - \hat{\varepsilon} - p(\varepsilon - \hat{\varepsilon} - \varepsilon\hat{\varepsilon})\}V_P\right], \quad (14)$$

$$V_P^{DD} = (1-\delta)(1+x) + \delta\left[(\varepsilon + \hat{\varepsilon} - \varepsilon\hat{\varepsilon})(1-p)V_R + \{1 - (\varepsilon + \hat{\varepsilon} - \varepsilon\hat{\varepsilon})(1-p)\}V_P\right]. \quad (15)$$

From Eqs. 8 and 9, we obtain

$$(1-\delta)x = \delta(2p-1)\hat{\varepsilon}(V_R - V_P). \quad (16)$$

Therefore, from Eqs. 8 and 12,

$$\begin{aligned} V_R - V_R^{DC} &= -(1-\delta)x + \delta(2p-1)(1-\hat{\varepsilon})(V_R - V_P) \\ &= \delta(2p-1)(1-2\hat{\varepsilon})(V_R - V_P) \geq 0, \end{aligned}$$

where the inequality follows from Eq. 7. From Eqs. 9 and 13, the same argument shows $V_R - V_R^{DD} \geq 0$. Therefore, from Eqs. 10, 14, and 16, we derive $V_P - V_P^{DC} = 0$. As well, Eqs. 11, 15, and 16 imply $V_P - V_P^{DD} = 0$.

Corollary 1. *Fix $M = 2$. For any δ , x , y , and p , if NTPD is an equilibrium, EV is an equilibrium and its payoff is greater than or equal to that of NTPD.*

This is straightforwardly derived from Theorem 1. The detailed proof is provided in Appendix A. Corollary 1 is somewhat negative. In fact, the NTPD's average payoff per market achieved in an equilibrium never exceeds that of EV. However, it is decreasing in the discount factor and is maximized with the lowest one so that the equilibrium condition is satisfied. With such a discount factor, if $x \geq y$, NTPD performs the same average payoff as EV, though we omit the calculation due to space constraints. In the next subsection, we show that the consequence can be reversed when considering more PDs than two.

3.2 M PDs

The following theorem identifies the equilibrium conditions and the average payoff for the case of M PDs.

Theorem 2 (NTPD for M PDs). *There exist ε and $\hat{\varepsilon}$ such that the NTPD strategy pair is a belief-free equilibrium if*

$$\begin{aligned} & \delta \left[x(1 - s^{M_A}) + s^{M_A-1} \left\{ M_B(p - s) - x(M_A - M_B)p - \frac{s^{M_B}(p - s)M_By}{p^{M_B} - s^{M_B}} \right\} \right] \\ & \geq x(1 + s^{M_A-1}M_B) \text{ and} \end{aligned} \quad (17)$$

$$\begin{aligned} & \delta \left[(p^{M_B} - s^{M_B}) \{ M_B(p - s) + (M_A - M_B)x(s - s^{M_A}) \} + M_By(p - s)(1 - s^{M_A} - s^{M_B}) \right] \\ & \geq M_Ax(p^{M_B} - s^{M_B}) + M_By(p - s) \end{aligned} \quad (18)$$

hold. The average payoff starting from R is

$$V_R = M - \frac{\delta s^{M_A}(p - s)(M - V_P) + (1 - \delta)(s - s^{M_A})M_Bx}{(p - s)\{1 - \delta(1 - s^{M_A})\}}, \quad (19)$$

$$\text{where } V_P = M_A + \frac{pM_Ax}{p-s} + \frac{s^{M_B}M_By}{p^{M_B} - s^{M_B}}.$$

Here, we refer to $1 - p$ as s for simplicity. Even though the proof is provided in Appendix B, it is basically the same as Theorem 1. Solving the system of value equations provides V_R , V_P , ε , and $\hat{\varepsilon}$. The conditions are derived from the incentive conditions and the feasibility conditions of ε and $\hat{\varepsilon}$. Note that, if p is sufficiently close to one, the equilibrium conditions are simplified to $\delta \approx 1$ and $M_B > M_Ax$.

The next corollary considers what happens if the numbers of PDs are sufficiently large.

Corollary 2. *Fix x , y , and p . Suppose both M_A and M_B are sufficiently large and satisfy*

$$M_A p x - (M_B - 1)\{2p - 1 - (1 - p)x\} \geq (1 - p)y. \quad (20)$$

Then if NTPD is an equilibrium for sufficiently large δ , EV is equilibrium, but its payoff is smaller than that of NTPD.

Proof. If NTPD is an equilibrium, Eqs. 17 and 18 hold. Hence, both of them evaluated at $\delta = 1$ hold with strict inequality. If M_A and M_B are sufficiently large, we have $s^{M_A} \rightarrow 0$ and $s^{M_B} \rightarrow 0$. Therefore, the two strict inequalities imply

$$M_B\{2p - 1 - (1 - p)x\} - M_A p x > 0.$$

This and Eq. 20 imply that $2p - 1 - (1 - p)(x + y) > 0$. It follows from Eq. 3 that EV is an equilibrium for sufficiently large δ .

Furthermore, we show that the NTPD's average payoff is greater than the EV's one. From Eq. 19, the payoff of NTPD when we let $s^{M_A} \rightarrow 0$ is

$$M - \frac{M_B(1 - p)x}{2p - 1} > M - \frac{M(1 - p)x}{2p - 1}.$$

The inequality follows from $M > M_B$. Since the right-hand side is the average payoff of EV (M times Eq. 4), the proof is complete.

The result is positive unlike Corollary 1. If the numbers of PDs are sufficiently large, NTPD achieves a greater payoff than EV. A question remains: how many PDs are required so that NTPD still outperforms EV? As we remarked at the end of Section 3.1, if we use the lowest discount factor so that NTPD is an equilibrium, it always outperforms EV, irrespective of the numbers of PDs. However, it is hard to analytically investigate cases when players are mildly patient ($\delta < 1$), because Corollary 2 cannot exactly be applied to the cases. Therefore, we provide a numerical analysis in Appendix C and show NTPD can yield better payoffs than EV for a specific number of PDs ($M = 6$).

4 Discussions

We have seen that NTPD sometimes outperforms EV. What aspects of NTPD contribute to such results? A key feature involves the nonlinearity of the transition probabilities. In fact that from state R to P does not depend on the outcome in A at all, as long as it contains at least one good signal. However, if all signals from A are bad, the transition probability sharply increases. In the former, that probability is $k\hat{\epsilon}$ where k is the number of bad signals in B . In the latter, it sharply increases to $1 - (M_B - k)\hat{\epsilon}$. This nonlinearity specifies the NTPD's first equilibrium condition in Eq. 17.

Why does this nonlinearity help? Suppose the other player is at state R , and consider how a player wants to play the PDs in A . Her incentive to play C or

D in one PD in A crucially depends on the probability of the event that all signals among the other PDs in A are bad. Only under that event, is her action in this PD pivotal. Naturally, the event is more likely when she defects among more PDs in A . Therefore, her temptation to defect in one PD in A is largest when she cooperates among all other PDs in A . Note that we apply a similar argument to this when we check the incentives. This observation implies that once an NTPD strategy prevents a player from defecting in one PD in A , it automatically ensures that the player has no incentive to defect in any number of PDs in A . Therefore, as long as we consider the NTPD strategies, we can effectively ignore all actions which defect among two or more PDs in A . This reduction in the number of incentive constraints is a key to the payoff improvement results brought about by NTPD. Conversely, NTPD never outperforms EV with two PDs as in Corporally 1 because this reduction is ineffective for proving Theorem 1.

This argument also reveals that the NTPD strategies must involve partial defection. Due to the nonlinearity, it is suboptimal to defect in all PDs, including the ones in B . Since the actions at states R and P must be both optimal in a belief-free equilibrium, full defection cannot be the action at state P .

Alternatively, we suspect a more complicated strategy, i.e., automata with more than two states, not to improve the payoff. Adding a new cooperation state is ineffective for achieving an equilibrium, since it gives a player a chance to exploit her opponent. Because the new state increases the likelihood that a player cooperates, even if the opponent defects, she is unlikely to shift a punishment state. Thus, he at a cooperation state can deviate to defect without being punished. Conversely, adding a punishment state may lead to an equilibrium. However, such a strategy inevitably decreases the payoff. Also, the idea of NTPD can be extended for games beyond PD, e.g., a game where a player has actions more than three. We can construct some similar strategy if a game has an efficient outcome, e.g., every player cooperates, and a Nash outcome, e.g., every player defects.

5 Conclusions

This paper identifies equilibria in repeated multimarket contact with a noisy signal. For the first time, we find the multimarket contact effect in the proposed class of strategies, NTPD, particularly when the number of PDs is large. In future work, we would like to improve NTPD and to characterize an optimal equilibrium strategy class.

References

1. Abreu, D., Pearce, D., Stacchetti, E.: Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58, 1041–1064 (1990)
2. Andersen, G., Conitzer, V.: Fast equilibrium computation for infinitely repeated games. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. pp. 53–59 (2013)

3. Bernheim, B.D., Whinston, M.D.: Multimarket Contact and Collusive Behavior . *RAND Journal of Economics* 21(1), 1–26 (1990)
4. Blum, A., Monsour, Y.: Learning, regret minimization, and equilibria. In: *Algorithmic game theory*, pp. 79–101. Cambridge University Press (2007)
5. Borgs, C., Chayes, J., Immorlica, N., Kalai, A.T., Mirrokni, V., Papadimitriou, C.: The myth of the folk theorem. *Games and Economic Behavior* 70(1), 34 – 43 (2010)
6. Burkov, A., Chaib-draa, B.: Repeated games for multiagent systems: A survey. *The Knowledge Engineering Review* pp. 1–30 (2013)
7. Chellappaw, R.K., Sambamurthy, V., Saraf, N.: Competing in crowded markets: Multimarket contact and the nature of competition in the enterprise systems software industry. *Information Systems Research* 21(3), 614–630 (2010)
8. Conitzer, V., Sandholm, T.: AWESOME: a general multiagent learning algorithm that converges in self-play and learns a best response against stationary opponents. *Machine Learning* 67(1), 23–43 (2007)
9. Doshi, P., Gmytrasiewicz, P.J.: On the Difficulty of Achieving Equilibrium in Interactive POMDPs. In: *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*. pp. 1131–1136 (2006)
10. Ely, J.C., Horner, J., Olszewski, W.: Belief-free equilibria in repeated games. *Econometrica* 73(2), 377–415 (2005)
11. Ely, J.C., Välimäki, J.: A robust folk theorem for the prisoner’s dilemma. *Journal of Economic Theory* 102(1), 84–105 (2002)
12. Hansen, E.A., Bernstein, D.S., Zilberstein, S.: Dynamic programming for partially observable stochastic games. In: *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*. pp. 709–715 (2004)
13. Kandori, M.: Repeated games. In: Durlauf, S.N., Blume, L.E. (eds.) *Game theory*, pp. 286–299. Palgrave Macmillan (2010)
14. Kobayashi, H., Ohta, K.: Optimal collusion under imperfect monitoring in multi-market contact. *Games and Economic Behavior* 76(2), 636–647 (2012)
15. Kreps, D.M., Wilson, R.: Sequential equilibria. *Econometrica* 50(4), 863–94 (1982)
16. Littman, M.L., Stone, P.: A polynomial-time Nash equilibrium algorithm for repeated games. *Decision Support Systems* 39(1), 55–66 (2005)
17. Mailath, G., Samuelson, L.: *Repeated Games and Reputation*. Oxford University Press (2006)
18. Mescheder, D., Tuyls, K., Kaisers, M.: POMDP opponent models for best response behavior. In: *Proceedings of the 23rd Benelux Conference on Artificial Intelligence* (2011)
19. Shoham, Y., Leyton-Brown, K.: *Learning and teaching*. In: *Multiagent systems: Algorithmic, Game-Theoretic, and Logical Foundations*, pp. 189–222. Cambridge University Press (2008)
20. Sugaya, T.: Folk theorem in repeated games with private monitoring (2015), revised and resubmitted to *Review of Economic Studies*
21. Tennenholtz, M., Zohar, A.: Learning equilibria in repeated congestion games. In: *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multi-Agent System (AAMAS)*. pp. 233–240 (2009)
22. Wunder, M., Kaisers, M., Yaros, J.R., Littman, M.: Using iterated reasoning to predict opponent strategies. In: *The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. pp. 593–600 (2011)
23. Yu, T., Cannella, A.: A comprehensive review of multimarket competition research. *Journal of Management* 39, 76–109 (2013)

Appendices

A Proof of Corollary 1

Let us first show that the NTPD's equilibrium conditions imply that of EV. From the equilibrium conditions of EV and NTPD for $M = 2$ (Eqs. 3 and 5), we require

$$\begin{aligned} & \delta[2p - 1 - (1 - p)(x + y) + \max\{x, y\}] \\ & \geq \delta[2p - 1 - (1 - p)y + p \max\{x, y\}] - x. \end{aligned}$$

Then we obtain

$$\delta[(1 - p)x - (1 - p) \max\{x, y\}] \leq x. \quad (21)$$

When $x \geq y$, Eq. 21 leads to $x \geq 0$. This always holds from the definition of $x > 0$. When $y > x$, Eq. 21 leads to $\delta[(1 - p)(y - x)] \geq -x$. This always holds because the lefthand side is always positive. Therefore, for any δ , x , y , and p , if NTPD is an equilibrium, EV is always an equilibrium.

Second, we show that the EV's average payoff is always greater than or equal to that of NTPD when it is an equilibrium. From the average payoffs of EV and NTPD for $M = 2$ (Eqs. 4 and 6), we require

$$2 - \frac{2(1 - p)x}{2p - 1} \geq 2 - \frac{\delta(1 - p)\{2p - 1 - px - (1 - p)y\}}{(2p - 1)(1 - \delta p)}.$$

Then we obtain

$$\delta[2p - 1 - (1 - p)y + px] \geq 2x. \quad (22)$$

Next, we show that the equilibrium condition of NTPD for $M = 2$ (Eq. 5 in the paper) implies Eq. 22. This requires

$$\begin{aligned} & \delta[2p - 1 - (1 - p)y + px] - x \\ & \geq \delta[2p - 1 - (1 - p)y + p \max\{x, y\}] - \max\{x, y\}. \end{aligned}$$

Since we then obtain $\max\{x, y\} \geq x$, this clearly holds. The proof is complete.

B Proof of Theorem 2

Define

$$\hat{\varepsilon} = \frac{x\{1 - \delta(1 - s^{M_A})\}}{\delta\left\{M_B(p - s) - x(M_A p + M_B s - M_B s^{M_A}) - \frac{s^{M_B}(p - s)M_B y}{p^{M_B} - s^{M_B}}\right\}}.$$

We claim that the first condition for the theorem implies

$$0 < \hat{\varepsilon} \leq \frac{s^{M_A - 1}}{1 + s^{M_A - 1} M_B} < 1. \quad (23)$$

To see that, note first that

$$\frac{s^{M_A-1}}{\hat{\varepsilon}(1+s^{M_A-1}M_B)} = \frac{\delta s^{M_A-1} \left\{ M_B(p-s) - x(M_A p + M_B s - M_B s^{M_A}) - \frac{s^{M_B}(p-s)M_B y}{p^{M_B} - s^{M_B}} \right\}}{\{1 - \delta(1-s^{M_A})\}(1+s^{M_A-1}M_B)x}. \quad (24)$$

Further, the first condition for the theorem implies

$$\begin{aligned} & 0 < \{1 - \delta(1-s^{M_A})\}(1+s^{M_A-1}M_B)x \\ & \leq \delta \left[s^{M_A-1} \left\{ M_B(p-s) - x(M_A - M_B)p - \frac{s^{M_B}(p-s)M_B y}{p^{M_B} - s^{M_B}} \right\} \right. \\ & \quad \left. + x(1-s^{M_A}) - (1-s^{M_A})(1+s^{M_A-1}M_B)x \right] \\ & = \delta s^{M_A-1} \left\{ M_B(p-s) - x(M_A p + M_B s - M_B s^{M_A}) - \frac{s^{M_B}(p-s)M_B y}{p^{M_B} - s^{M_B}} \right\}. \end{aligned}$$

Substituting this into Eq. 24 yields Eq.23.

Next, define

$$\varepsilon = \frac{M_B y(p-s)\hat{\varepsilon}}{(1-M_A\hat{\varepsilon})(p^{M_B} - s^{M_B})x}.$$

The second condition for the theorem implies

$$\begin{aligned} & \{M_A x(p^{M_B} - s^{M_B}) + M_B y(p-s)\} \{1 - \delta(1-s^{M_A})\} \\ & \leq \delta \left[(p^{M_B} - s^{M_B}) \{M_B(p-s) + (M_A - M_B)x(s - s^{M_A})\} \right. \\ & \quad \left. + M_B y(p-s)(1-s^{M_A} - s^{M_B}) - (1-s^{M_A}) \{M_A x(p^{M_B} - s^{M_B}) + M_B y(p-s)\} \right] \\ & = \delta \left[(p^{M_B} - s^{M_B}) \{M_B(p-s) - x(M_A p + M_B s - M_B s^{M_A})\} - (p-s)s^{M_B} M_B y \right], \end{aligned}$$

which is equivalent to

$$\{M_A x(p^{M_B} - s^{M_B}) + M_B y(p-s)\} \hat{\varepsilon} \leq (p^{M_B} - s^{M_B})x.$$

Rearranging, we obtain

$$0 < M_B y(p-s)\hat{\varepsilon} \leq (1-M_A\hat{\varepsilon})(p^{M_B} - s^{M_B})x,$$

which verifies that $0 < \varepsilon \leq 1$.

Let us now consider the NTPD strategy with ε and $\hat{\varepsilon}$ defined as above. First, we have the following value equations.

$$V_{RR} = (1-\delta)M + \delta V_{RR} - \delta(V_{RR} - V_{RP}) \{M_B s \hat{\varepsilon} + s^{M_A}(1 - M_B \hat{\varepsilon})\}, \quad (25)$$

$$\begin{aligned} V_{RP} &= (1-\delta)(M_A - M_B y) + \delta V_{RP} + \delta(V_{RR} - V_{RP}) \{M_A p \hat{\varepsilon} + p^{M_B} \varepsilon(1 - M_A \hat{\varepsilon})\} \\ &= (1-\delta)(M_A - M_B y) + \delta V_{RP} + \delta(V_{RR} - V_{RP}) \left\{ M_A p + \frac{p^{M_B} M_B y(p-s)}{(p^{M_B} - s^{M_B})x} \right\} \hat{\varepsilon}, \end{aligned} \quad (26)$$

where Eq. 26 follows from the definition of ε . It follows from Eqs. 25 and 26 that

$$(1 - \delta)M_B(1 + y) = (V_{RR} - V_{RP}) \left[1 - \delta + \delta s^{M_A} + \left\{ M_B s - M_B s^{M_A} + M_A p + \frac{p^{M_B} M_B y (p - s)}{(p^{M_B} - s^{M_B})x} \right\} \delta \hat{\varepsilon} \right]. \quad (27)$$

Note that from the definition of $\hat{\varepsilon}$,

$$1 - \delta(1 - s^{M_A}) = \delta \hat{\varepsilon} \left\{ \frac{M_B(p - s)}{x} - (M_A p + M_B s - M_B s^{M_A}) - \frac{s^{M_B}(p - s)M_B y}{(p^{M_B} - s^{M_B})x} \right\}. \quad (28)$$

Let us substitute this into Eq. 27.

$$(1 - \delta)M_B(1 + y) = \delta \hat{\varepsilon} (V_{RR} - V_{RP}) \left\{ \frac{M_B(p - s)}{x} - \frac{s^{M_B}(p - s)M_B y}{(p^{M_B} - s^{M_B})x} + \frac{p^{M_B} M_B y (p - s)}{(p^{M_B} - s^{M_B})x} \right\} \\ = \frac{\delta \hat{\varepsilon} M_B(1 + y)(p - s)}{x} (V_{RR} - V_{RP}),$$

from which we have

$$(1 - \delta)x = \delta \hat{\varepsilon}(p - s)(V_{RR} - V_{RP}). \quad (29)$$

Let us define

$$V_P \equiv M_A + \frac{pM_A x}{p - s} + \frac{s^{M_B} M_B y}{p^{M_B} - s^{M_B}}, \\ V_R \equiv M - \frac{\delta s^{M_A}(p - s)(M - V_P) + (1 - \delta)(s - s^{M_A})M_B x}{(p - s)\{1 - \delta(1 - s^{M_A})\}}.$$

We claim that $V_{RR} = V_R$ and $V_{RP} = V_P$. First, it follows from Eq. 26 that

$$V_{RP} = M_A - M_B y + \frac{\delta}{1 - \delta} (V_{RR} - V_{RP}) \left\{ M_A p + \frac{p^{M_B} M_B y (p - s)}{(p^{M_B} - s^{M_B})x} \right\} \hat{\varepsilon},$$

and substituting Eq. 29 yields $V_{RP} = V_P$. Next, note that

$$\hat{\varepsilon} = \frac{x\{1 - \delta(1 - s^{M_A})\}}{\delta\{(p - s)(M - V_P) - (s - s^{M_A})M_B x\}}. \quad (30)$$

From Eqs. 25, 29 and 30, we obtain

$$M - V_{RR} = \frac{\delta}{1 - \delta} (V_{RR} - V_{RP}) \{M_B(s - s^{M_A})\hat{\varepsilon} + s^{M_A}\} \\ = \frac{M_B x(s - s^{M_A})}{p - s} + \frac{x s^{M_A}}{(p - s)\hat{\varepsilon}} \\ = \frac{\{1 - \delta(1 - s^{M_A})\}M_B x(s - s^{M_A}) + \delta s^{M_A}\{(p - s)(M - V_P) - (s - s^{M_A})M_B x\}}{(p - s)\{1 - \delta(1 - s^{M_A})\}} \\ = \frac{\delta s^{M_A}(p - s)(M - V_P) + (1 - \delta)(s - s^{M_A})M_B x}{(p - s)\{1 - \delta(1 - s^{M_A})\}} = M - V_R.$$

Therefore, $V_{RR} = V_R$.

We also have the following value equations.

$$V_{PR} = (1 - \delta)(M + M_B x) + \delta V_{PR} - \delta(V_{PR} - V_{PP}) \{M_B p \hat{\varepsilon} + s^{M_A}(1 - M_B \hat{\varepsilon})\}, \quad (31)$$

$$\begin{aligned} V_{PP} &= (1 - \delta)M_A + \delta V_{PP} + \delta(V_{PR} - V_{PP}) \{M_A p \hat{\varepsilon} + s^{M_B} \varepsilon(1 - M_A \hat{\varepsilon})\} \\ &= (1 - \delta)M_A + \delta V_{PP} + \delta(V_{PR} - V_{PP}) \left\{ M_A p + \frac{s^{M_B} M_B y(p - s)}{(p^{M_B} - s^{M_B})x} \right\} \hat{\varepsilon}, \end{aligned} \quad (32)$$

where Eq. 32 follows from the definition of ε . It follows from Eqs. 31 and 32 that

$$\begin{aligned} &(1 - \delta)M_B(1 + x) \\ &= (V_{PR} - V_{PP}) \left[1 - \delta + \delta s^{M_A} + \left\{ M_B(p - s^{M_A}) + M_A p + \frac{s^{M_B} M_B y(p - s)}{(p^{M_B} - s^{M_B})x} \right\} \delta \hat{\varepsilon} \right] \\ &= (V_{PR} - V_{PP}) \delta \hat{\varepsilon} \left\{ \frac{M_B(p - s)}{x} - M_B s + M_B p \right\} = (V_{PR} - V_{PP}) \delta \hat{\varepsilon} \frac{x + 1}{x} M_B(p - s), \end{aligned}$$

where the second equality is due to Eq. 28. This is equivalent to

$$(1 - \delta)x = \delta \hat{\varepsilon}(p - s)(V_{PR} - V_{PP}). \quad (33)$$

From Eq. 32, we have

$$V_{PP} = M_A + \frac{\delta}{1 - \delta}(V_{PR} - V_{PP}) \left\{ M_A p + \frac{s^{M_B} M_B y(p - s)}{(p^{M_B} - s^{M_B})x} \right\} \hat{\varepsilon},$$

and substituting Eq. 33 proves $V_{PP} = V_P$. Comparing Eqs. 29 and 33, and using $V_{PP} = V_{RP}$, we obtain $V_{PR} = V_{RR} = V_R$.

It suffices to verify that (i) V_R is a player's best response payoff when the other player is at state R , and (ii) V_P is a player's best response payoff when the other player is at state P . To this end, let $V(d_A, d_B, s)$ be a player's continuation payoff when he chooses D in some d_A PDs in A and in some d_B PDs in B and then conforms to the NTPD strategy from the next period on, given that the other player is at state s and conforms to the NTPD strategy. The proof is complete if we show that $V(0, 0, s) \geq V(d_A, d_B, s)$ for any d_A , any d_B , and any s .

First, note that

$$\begin{aligned} V(d_A, d_B, R) &= (1 - \delta)(M + d_A x + d_B x) + \delta V_R \\ &\quad - \delta(V_R - V_P) \left[\{ (M_B - d_B)s + d_B p \} \hat{\varepsilon} + p^{d_A} s^{M_A - d_A} (1 - M_B \hat{\varepsilon}) \right]. \end{aligned}$$

This is linear in d_B , and its slope is

$$(1 - \delta)x - \delta(V_R - V_P)(p - s)\hat{\varepsilon} = 0, \quad (34)$$

where the equality follows from Eq. 29 and $V_{RR} - V_{RP} = V_R - V_P$. Hence, $V(d_A, d_B, R)$ does not depend on d_B . Note also that for any d_A ,

$$\begin{aligned} & V(d_A, 0, R) - V(d_A + 1, 0, R) \\ &= -(1 - \delta)x + \delta(V_R - V_P)(p^{d_A+1}s^{M_A-d_A-1} - p^{d_A}s^{M_A-d_A})(1 - M_B\hat{\varepsilon}) \\ &= \delta(V_R - V_P)(p - s)\{p^{d_A}s^{M_A-d_A-1}(1 - M_B\hat{\varepsilon}) - \hat{\varepsilon}\}, \end{aligned}$$

where the last equality is due to Eq. 34. Since $p > s$, Eq. 23 implies

$$p^{d_A}s^{M_A-d_A-1}(1 - M_B\hat{\varepsilon}) - \hat{\varepsilon} \geq s^{M_A-1}(1 - M_B\hat{\varepsilon}) - \hat{\varepsilon} \geq 0.$$

Therefore, $V(d_A, 0, R) \geq V(d_A + 1, 0, R)$ for any d_A . From these, we obtain

$$V(d_A, d_B, R) = V(d_A, 0, R) = \sum_{d=1}^{d_A} \{V(d, 0, R) - V(d-1, 0, R)\} + V(0, 0, R) \leq V(0, 0, R)$$

for any d_A and any d_B , as desired.

Finally, note that

$$\begin{aligned} V(d_A, d_B, P) &= (1 - \delta)\{M_A + d_Ax - (M_B - d_B)y\} + \delta V_P \\ &\quad + \delta(V_R - V_P)\left[\{(M_A - d_A)p + d_As\}\hat{\varepsilon} + s^{d_B}p^{M_B-d_B}\varepsilon(1 - M_A\hat{\varepsilon})\right]. \end{aligned}$$

This is linear in d_A , and its slope is zero by Eq. 34. Hence, $V(d_A, d_B, P)$ does not depend on d_A . Note also that for any d_B ,

$$\begin{aligned} & V(0, 0, P) - V(0, d_B, P) \\ &= -(1 - \delta)d_By + \delta(V_R - V_P)\varepsilon(1 - M_A\hat{\varepsilon})(p^{M_B} - s^{d_B}p^{M_B-d_B}) \\ &= \frac{y}{x}\left\{-(1 - \delta)d_Bx + \delta(V_R - V_P)\frac{M_B(p-s)\hat{\varepsilon}}{p^{M_B} - s^{M_B}}(p^{M_B} - s^{d_B}p^{M_B-d_B})\right\} \\ &= \frac{y}{x}\delta(V_R - V_P)(p-s)\hat{\varepsilon}\left\{-d_B + \frac{M_B}{p^{M_B} - s^{M_B}}(p^{M_B} - s^{d_B}p^{M_B-d_B})\right\}, \end{aligned}$$

where the second and third equalities follow from the definition of ε and Eq. 34, respectively. Since $p > s$, this is concave in d_B . Further, it attains the same value at $d_B = 0$ and $d_B = M_B$, and concavity therefore implies that $V(0, 0, P) \geq V(0, d_B, P)$ for any d_B . Since

$$V(d_A, d_B, P) = V(0, d_B, P) \leq V(0, 0, P)$$

for any d_A and any d_B , the proof is complete.

C Numerical analysis

This section numerically evaluates NTPD. Throughout this section, we fix the stage game payoffs at $x = y = 0.1$ and assume a player in NTPD defects in the

half number of the PDs at state P .⁷ Figure 3 examines the average payoffs of NTPD and EV with six PDs ($M = 6$ and $M_B = M_A = 3$). The x-axis indicates the correctness of signals p while the y-axis indicates the average payoffs per PD. We plot the NTPD's payoffs with $\delta \in \{0.7, 0.8, 0.9\}$. We show only the EV's payoff with $\delta = 0.7$ because the payoff is independent of δ and, for lower δ , only the lower limit of p decreases.

EV is an equilibrium in the range of $p \in [0.56, 0.99]$ and the payoffs increase as p does. EV is clearly an equilibrium in a wider range than NTPD for any discount factor: NTPD is an equilibrium when $p \in [0.64, 0.83]$ for $\delta = 0.7$, $p \in [0.60, 0.88]$ for $\delta = 0.8$, and $p \in [0.58, 0.92]$ for $\delta = 0.9$. Such an upper bound of p that NTPD is an equilibrium exists for a given δ because Eq. 17 requires δ to be high when p is high. A lower bound exists because Eq. 17 unlikely holds simply because the lefthand side is small and the righthand side is large when p is low. The NTPD's average payoffs are basically outperformed by the EV's one with $\delta = 0.9$. However, as δ is lowered, the NTPD's payoffs gradually increase. When $\delta = 0.7$, NTPD always outperforms EV with a maximum 4.75 % increase and a minimum 0.87 % increase. In addition, we confirm that a further low discount factor admittedly magnifies the difference as do larger numbers of PDs ($M > 6$) and that, if $M > 6$, there exists some parameter setting such that, if NTPD is an equilibrium, EV is an equilibrium and its payoff is greater than that of NTPD.

It must be emphasized that whether this improvement is subtle depends on how much a player actually values the payoffs. For example, if she is a president of a firm and the sales are a million dollars per month, a few percent increase definitely deserves her attention. This is also a reason why we assume that the gain from defection x and the loss caused by the opponent's defection y are small. In any case, besides the case of large M_A and M_B covered by Corollary 2, we could construct NTPD beating EV when the number of markets is relatively small.

The next question addresses whether NTPD achieves an optimal payoff among possible equilibrium strategies. We examine how efficient the NTPD's transition from state R . Under perfect or public monitoring, it is known that a player's equilibrium payoff vector can be computed independently from the opponent's one. Dynamic programming can derive the bounds of the equilibrium payoff of each player, i.e., a *self-generation* set [1]. It is guaranteed that an equilibrium strategy with a payoff vector in the set exists.

Under private monitoring, this is generally impossible. However, if an equilibrium is belief-free, since it satisfies an *exchangeability* property, the payoff set has a product structure [10]. Thus, given a continuation payoff starting from state P , the upper bound of one from state R is computed by the following linear programming:

⁷ We confirm that this fraction performs best when NTPD can be an equilibrium, irrespective of the total number of PDs M .

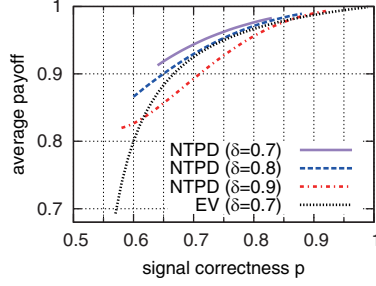


Fig. 3: Average payoffs

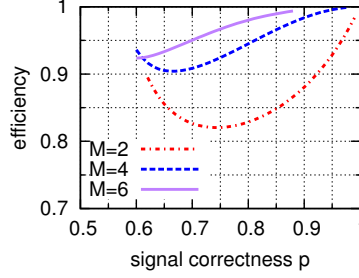


Fig. 4: Efficiency

$$\begin{aligned}
& \max v_i \\
& \text{s.t. } v_i \geq (1 - \delta) \sum_{a_j} \alpha(a_j) g_i(a_i, a_j) \\
& \quad + \delta \left[\sum_{\omega_j} \sum_{a_j} o_j(\omega_j \mid a_i, a_j) \alpha(a_j) z_i(a_j, \omega_j) \right] \\
& \text{for each } a_i \in \{C, D\}^M, \text{ with equality for each } a_i \in \{f_i(R), f_i(P)\}, \\
& z_i(a_j, \omega_j) \leq v_i \text{ for all } a_j \in \{C, D\}^M \text{ and } \omega_j \in \{g, b\}^M, \\
& z_i(a_j, \omega_j) \geq V_P \text{ for all } a_j \in \{C, D\}^M \text{ and } \omega_j \in \{g, b\}^M, \\
& \sum_{a_j \in \{C, D\}^M} \alpha(a_j) = 1, \text{ and } \alpha(a_j) \in [0, 1] \text{ for all } a_j \in \{C, D\}^M.
\end{aligned}$$

Note that $i \neq j$, if $i = 1, j = 2$, otherwise $j = 1$. The first constraints are incentive constraints: if a player obtains v_i when she employs a strategy, deviating from the strategy is not profitable. $f_i(R)$ or $f_i(P)$ indicates actions specified at state R or P . The decision variables are player j 's mixed action $\alpha(a_j)$ and the product of the mixed action and i 's continuation payoff that is dependent on j 's current action and observation $\alpha(a_j)z_i(a_j, \omega_j)$. If i 's action is specified at a state, the constraints must be defined with equality to satisfy belief-freeness. The next two specify the feasible bounds of the equilibrium payoffs. V_P is given a priori, i.e., the NTPD's average payoff starting from state P . Thus, the solution implies maximizing the continuation payoff starting from state R . The last two define the feasibility for $\alpha(a_j)$.

Figure 4 illustrates the NTPD's efficiency, i.e., the ratio of the NTPD's average payoff to the optimal payoff. The x- and y-axes indicate the ratio and signal correctness p , respectively. We fix the discount factor at $\delta = 0.8$. When $M = 2$ or 4, the efficiency decreases once, increases in p , and reaches 0.99 and 0.98. When $M = 6$, it reaches 0.99, though NTPD is no longer an equilibrium when p exceeds 0.88. Observe that, when $M = 6$, NTPD performs at over 90% efficiency.⁸ There is a room for improving the NTPD's transitions from R . Whether we can further improve the transition is our immediate future work.

⁸ For the larger number of PDs ($M > 6$), we are unable to solve the linear programming by CPLEX 12.5 on Windows 8.1 Pro x64 PC with a Core i7-3960X and 32GB RAM.